# SOME PROBLEMS IN ANIMAL EXPERIMENTATION

by

Walter T. Federer

BU-810-M*

March 1983

## Abstract and Summary

A statistical consultant is faced with many unsolved problems in advising on the design and analysis of animal experiments. Some of the problems arise because of inadequate or inappropriate statistical procedures. Statistical textbooks may treat situations which do not hold in animal experimentation. Teachers of statistics classes omit many published techniques that could be useful to the animal scientist. One noteable example is the treatment of a statistical analysis for a Latin square design. An analysis of the type given by C. P. Cox (1958) would be more suitable than textbook analyses for a portion of animal experiments using a Latin square design.

Repeated measures designs are rarely discussed and little understood. There is much more to repeated measures designs than construction or using linear model theory. The nature of the treatment effect and its duration are vitally important in making intelligent use of such designs. Also, statistical optimality may not be the only usable criterion in selecting a repeated measures design. Two particular examples are discussed with one raising problems of conduct of the experiment and appropriateness of the model.

An example in which the nature of the experiment dictates the type of experiment design with no alternatives is given. The nature of the treatments dictate the dam as the whole plot unit and the sex within a litter as the split plot treatment. Some aspects of analyzing this experiment are discussed.

In the last section of the paper a variety of problems encountered are discussed briefly. Among them are the unconditional analyses for unbalanced classifications, covariance, multivariate analyses, measurements, outliers, data bases, animal breeding and variance components, recommended levels, Waller-Duncan procedure, response plateaus, and lactation curves.

---

* BU-810-M in the Biometrics Unit Series, Cornell University, Ithaca, New York.

## 1. Introduction

During the course of statistical consulting on problems in animal science, a number of situations have been encountered wherein no statistical procedures were available or an incorrect or inappropriate procedure was being used. The first example has to do with statistical analyses for row-column designs, starting with a Latin square design and some statistical analyses. Repeated measures row-column designs are considered next. Then, a row-column design with problems is discussed. An example of a split-plot design for which it is impossible to have another design is discussed. Finally, a number of areas of concern (problems) are raised. Most of these are unsolved problems which need answers for present animal science research needs.

## 2. Row-Column Designs – Latin Square

A frequently used design in animal experimentation for comparing a set of v treatments is a classical Latin square design, with animals as the columns and time periods as the rows of the Latin square. For purposes of statistical analysis the following standard textbook model and analysis are used for a Latin square of order v. The response model, implicitly or explicitly, used is

$$Y_{ihi} = \mu + \rho_h + \gamma_i + \tau_j + \epsilon_{hij} \ ,$$

where $\mu$ is an overall mean effect, $\rho_h$ is the $h^{th}$ row (period) effect, $\gamma_i$ is the $i^{th}$ column (animal) effect, $\tau_j$ is the $j^{th}$ treatment effect, and $\epsilon_{hij}$ are random error deviations which are NIID$(0,\sigma^2_\epsilon)$. The analysis of variance table used is:

| Source of variation | Degrees of freedom | Sum of squares | Mean square |
|---|---|---|---|
| Total | $v^2$ | | |
| Correction for mean | 1 | | |
| Periods | v-1 | | |
| Animals | v-1 | | |
| Animals × periods | $(v-1)^2$ | | |
| Treatments | v-1 | | |
| "Error" | (v-1)(v-2) | | |

Occasionally, a one-degree of freedom sum of squares for non-additivity is computed using a response model of the form

$$Y_{hij} = \frac{\mu_{h..}\mu_{i..}\mu_{..j}}{\mu^2_{...}} + \epsilon^{*}_{hij}$$

$$= \mu + \rho_h + \gamma_i + \tau_j + \frac{1}{\mu^2_{...}}(\rho_h\gamma_i + \rho_h\tau_j + \gamma_i\tau_j)$$

$$+ \frac{1}{\mu^2_{...}}\rho_h\gamma_i\tau_j + \epsilon^{*}_{hij} \quad ,$$

where $\mu_{h..}$ is the $h^{th}$ row mean, $\mu_{.i.}$ is the $i^{th}$ column mean, $\mu_{..j}$ is the $j^{th}$ treatment mean, $\rho_h = \mu_{h..} - \mu_{...}$, $\gamma_i = \mu_{.i.} - \mu_{...}$, $\tau_j = \mu_{..j} - \mu_{...}$ and the other symbols are as defined above. Tukey's one-degree-of-freedom for non-additivity is computed as

$$\frac{\left[\sum_{h,i,j} \hat{\epsilon}_{hij}(\hat{\epsilon}'_{hij} - \hat{\epsilon}_{hij})\right]^2}{\sum_{h,i,j}(\hat{\epsilon}'_{hij} - \hat{\epsilon}_{hij})^2}$$

where $\hat{\epsilon}'_{hij} = \hat{\epsilon}^{*}_{hij} + \frac{1}{\hat{\mu}^2}\hat{\rho}_h\hat{\gamma}_i\hat{\tau}_j$ and $\hat{\epsilon}_{hij} = Y_{hij} - \bar{y}_{h..} - \bar{y}_{.i.} - \bar{y}_{..j} + 2\bar{y}_{...}$ . One could have used $\hat{\epsilon}^{*}_{hij} = Y_{hij} - \bar{y}_{h..}\bar{y}_{.i.}\bar{y}_{..j}/\bar{y}^2_{...}$ in place of $\hat{\epsilon}'_{hij}$ to obtain a slightly different test for nonadditivity, and which is computationally simpler than Tukey's.

It is highly likely that neither of the above response model equations and analyses of variance are appropriate for experiments of this type. The animals may be on different parts of a growth curve, on different parts of a lactation curve, or may be responding differently through time. The gradients over time for the animal could be different. The previous two response model equations assume they are the same. Cox (1958) has given a response model and an analysis of variance which considers that there are differential gradients over time for each animal. The response equation is:

$$Y_{ij} = \mu + \gamma_i + \beta_i a_{ij} + \tau_j + \epsilon_{ij} \quad ,$$

where $\mu + \gamma_i$ is the intercept and $\beta_i$ is the linear regression for animal i, the $a_{ij}$ are the elements of the design matrix, and the other components are as defined above. The partitioning of the degrees of freedom in the analysis of variance is:

| Source of variation | Degrees of freedom |
|---|---|
| Total | $v^2$ |
| Correction for mean | 1 |
| Animals | v-1 |
| Regressions for animals (ignoring treatments) | v |
| Treatments (eliminating individual animal regression) | v-1 |
| Error | (v-1)(v-2)-1 |

It should be noted that one could, as Cox (1958) did, add additional polynomial terms to the above response model equation.

To illustrate how effective Cox's (1958) analysis can be, he used four cows in their $17^{th}$, $11^{th}$, $15^{th}$, and $7^{th}$ week of lactation. The treatments are dummy treatments, and hence the treatment mean square and error mean squares should estimate the same parameter if the model is correct. The ANOVAs for his data are:

| Conventional analysis | | | Differential gradient analysis | | |
|---|---|---|---|---|---|
| Source | d.f. | m.s. | Source | d.f. | m.s. |
| Cows | 3 | 8525 | Cows | 3 | 8525 |
| Periods | 3 | 399 | Regressions (ign. tr.) | 4 | 553 |
| Treatments | 3 | 86 | Treatments (elim. reg.) | 3 | 36 |
| Remainder | 6 | 177 | Remainder | 5 | 40 |

With the conventional analysis, the remainder mean square was double the treatment mean square, whereas they were essentially equal, as they should be, in the Cox (1958) analysis. The "error" mean square from the conventional analysis was

over four times larger than the "error" from the Cox (1958) analysis. This illustrates that dramatic differences can be obtained using different analyses on the same data. Inappropriateness of a model can result in incorrect inferences.

It may be noted that a Latin square design laid out in a pasture experiment, for example, may have differential gradients in each row and in each column. Gradients which are not perpendicular to rows and to columns could give rise to such a situation.

## 3. Row-Column Designs – Carry-over Effects

In any experiment involving repeated measures through time on the same sampling unit many types of treatment response are possible. Some of these are illustrated in Figure 1. The first one, 1a, depicts an immediate treatment effect, which ceases once the treatment is discontinued. 1b depicts a treatment response that takes some time for treatment effect to be asserted, for example, most diets; the treatment effect reaches a maximum at time $t_1$, the end of the period, and ceases completely once the treatment is removed. In 1c, the treatment effect does not reach a maximum until after the treatment is discontinued at time $t_1$. Its effect continues over the next two periods. In order to measure the maximum response to a treatment, the period should have been longer than $t_1 - t_0$. If it had been one-third longer, and if the response had been measured at this time, the full effect of the treatment would have been measured. Its carry-over effect would not be present if a measurement was taken at time $t_3$. This demonstrates the necessity of selecting the correct length of a treatment period. The nature of the treatments and of their responses determine length of period. It is necessary to know something about effect of treatments on responses in order to select an appropriate experiment design and an appropriate treatment period.

There are many types of treatment effects in a repeated measures situation. Some of these follow:

A direct effect is the effect of treatment during period in which it was applied.

A one-period carry-over effect is the effect of the treatment in the period following the period of application (Figure 1c).

A two-period carry-over effect is the effect of the treatment in the second period following the period of application (Figure 1c).

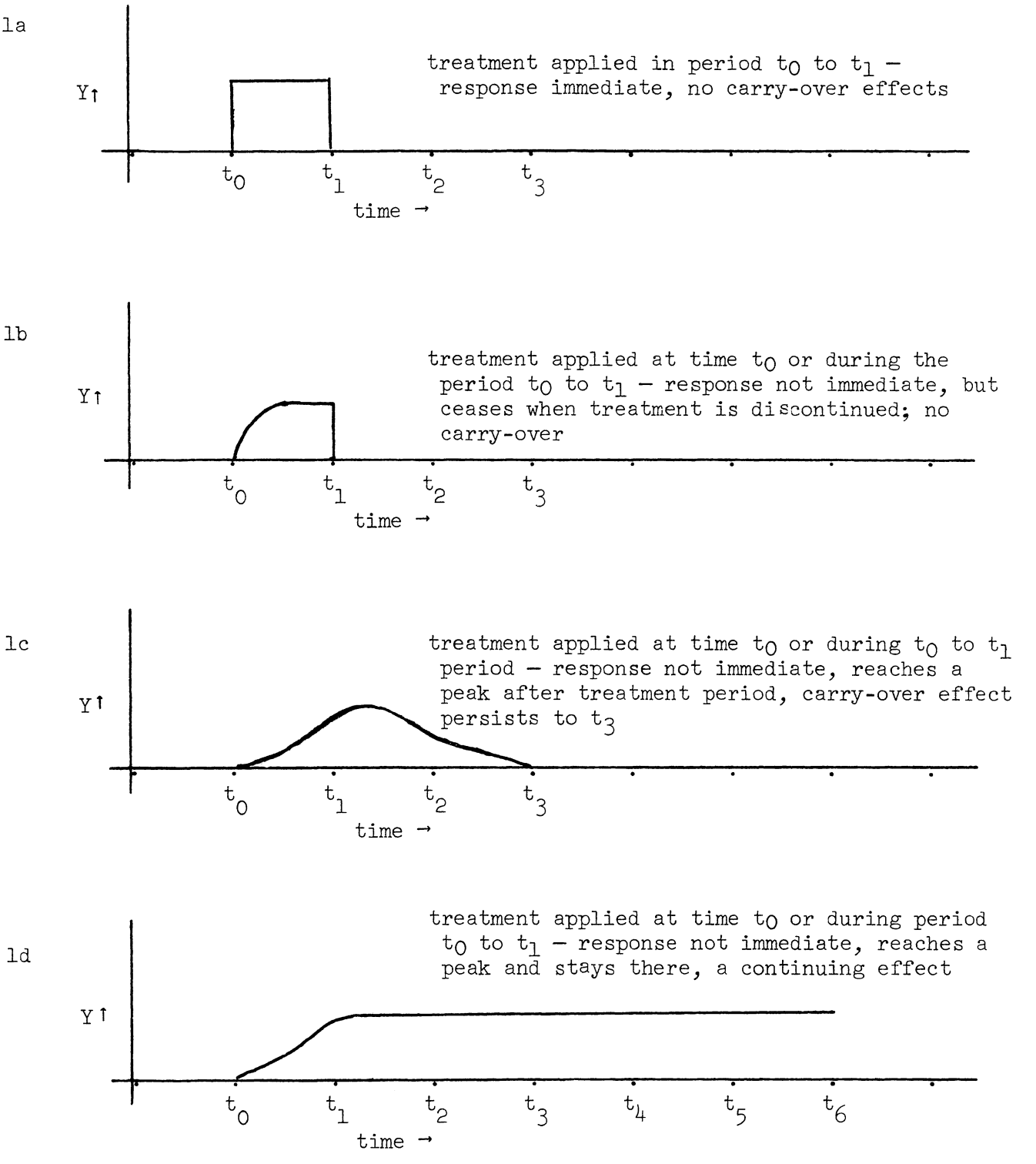


Figure 1. Some types of treatment responses, Y.

A three-period carry-over effect is the effect of the treatment in the third period following the period of applications, etc.

A continuing effect is the undiminishing effect of treatment over time (Figure 1d). This could happen if the treatment cures an animal, makes it sick, kills it, etc.

A cumulative effect is the sum of the direct plus all carry-over effects. This has also been called a permanent effect.

Not all row-column designs will allow estimation of carry-over effects. For example, the following Latin square plans do not allow estimation of carry-over effects:

| Rows | Columns 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | A | B | C | D |
| 2 | D | A | B | C |
| 3 | C | D | A | B |
| 4 | B | C | D | A |

| Rows | Columns 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | A | B | C | D | E | F |
| 2 | F | A | B | C | D | E |
| 3 | E | F | A | B | C | D |
| 4 | D | E | F | A | B | C |
| 5 | C | D | E | F | A | B |
| 6 | B | C | D | E | F | A |

However, the following two do:

| | | | |
|---|---|---|---|
| A | B | C | D |
| B | C | D | A |
| D | A | B | C |
| C | D | A | B |

| | | | | | |
|---|---|---|---|---|---|
| A | B | C | D | E | F |
| B | C | D | E | F | A |
| F | A | B | C | D | E |
| C | D | E | F | A | B |
| E | F | A | B | C | D |
| D | E | F | A | B | C |

The second set of squares is obtained by changing the order of the rows in the first. If the row order for the Latin square of order four is changed to 1 4 2 3, the second square emerges. Likewise, changing the row order for the Latin square of order six to 1 6 2 5 3 4 results in the second square of order six. In the first square of order four, B precedes A and D follows A in every column. In the

second square A is preceded and is followed by every other treatment. This arrangement holds for all treatments. Such arrangements allow estimation of carry-over effects, whereas the former set does not. These are easy to construct for all Latin squares of even order. Simply start with a cyclic square and push the first n/2 rows one row apart. Fill in the $n^{th}$ row with the $(n+2)/2^{nd}$ row of the original square, and proceed back up the square until the $n^{th}$ row of the original square falls in the second row of the desired square, i.e., the row order of the original square is changed to $1,v,2,(v-1),3,(v-2),\cdots,v/2,(v+2)/2$.

For v odd, it is necessary to have at least a pair of orthogonal Latin squares in order to have a design balanced for residual effects for v = 2, 5, and 7. For some larger values of v, one square has been found. Two Latin squares that are orthogonal have every treatment number in one square appearing once with any given treatment number of the second square. A pair of orthogonal squares for v = 3 and 5 follows:

$L_1(3)$

| | | |
|---|---|---|
| A | B | C |
| B | C | A |
| C | A | B |

$L_2(3)$

| | | |
|---|---|---|
| A | B | C |
| C | A | B |
| B | C | A |

$L_1(3) \perp L_2(3)$

$L_1(5)$

| | | | | |
|---|---|---|---|---|
| A | B | C | D | E |
| B | C | D | E | A |
| C | D | E | A | B |
| D | E | A | B | C |
| E | A | B | C | D |

$L_2(5)$

| | | | | |
|---|---|---|---|---|
| A | B | C | D | E |
| C | D | E | A | B |
| E | A | B | C | D |
| B | C | D | E | A |
| D | E | A | B | C |

$L_1(5) \perp L_2(5)$

Note that the main right diagonal of $L_1(3)$ and of $L_1(5)$ is the first column of $L_2(3)$ and of $L_2(5)$, respectively. The remaining letters in a row of $L_2(3)$ or $L_2(5)$ are obtained by writing the letters in order in a circular fashion. This construction procedure produces a pair of orthogonal Latin squares of odd order. For v a prime, continued application of the procedure produces a set of v-1 pairwise orthogonal Latin squares.

The following repeated measures design was constructed for Dr. P. Van Soest, Department of Animal Science, Cornell University, to study the effect of different types and amounts of fiber in diets on 24 healthy young males for several characteristics. Since the length of treatment period (the time on a diet) had to be determined after the experiment was started, the experiment

design had to be capable of allowing estimation of direct and residual effects with a flexible number of periods. Originally, it was thought that it would require a period of 20-25 days, but it became evident that less than two weeks were required to obtain stabilization for measurements on the various characteristics. Hence, the treatment period was set at two weeks, but some measurements were taken weekly even though the treatments did not change more often than every two weeks. The following design was constructed:

| | Boys | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Period | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | A | B | C | D | A | B | C | D | A | B | C | D |
| 2 | B | A | D | C | D | C | B | A | C | D | A | B |
| 3 | C | D | A | B | B | A | D | C | D | C | B | A |
| 4 | D | C | B | A | C | D | A | B | B | A | D | C |
| 5 | A | B | C | D | A | B | C | D | A | B | C | D |
| 6 | B | A | D | C | D | C | B | A | C | D | A | B |

| | Boys | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Period | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 1 | B | C | D | A | C | D | A | B | D | A | B | C |
| 2 | A | A | A | B | B | B | C | C | C | D | D | D |
| 3 | A | A | A | B | B | B | C | C | C | D | D | D |
| 4 | A | A | A | B | B | B | C | C | C | D | D | D |
| 5 | A | A | A | B | B | B | C | C | C | D | D | D |
| 6 | A | A | A | B | B | B | C | C | C | D | D | D |

There are many items associated with the conduct of the experiment which cause difficulties in the analysis, but the main point here is, why two designs? The design for boys 1-12 (selected at random) is a variance-optimal design, whereas the design for boys 13-24, although balanced for carry-over effects, is variance-minimal, i.e., it has the highest variance for differences between treatment effects among all designs balanced for residual effects. The question then is, why was it used? The answer is that Dr. Van Soest knew that he had to leave boys on a diet for a relatively long period of time, say 45-60 days, in order to convince fellow nutritionists to believe the results. If the results for boys

1-12 were similar to those obtained for boys 13-24, nutritionists would believe the results. Hence, the design for boys 1-12 is statistically optimal, whereas the one for boys 13-24 is nutritionally optimal. This illustrates that more than statistical criteria may be involved in choosing an "optimal" design.

## 4. Row-Column Designs – An Example with Problems

An experimenter wished to compare $2^4$ treatments involving all possible combinations of four factors at two levels each. He had four cows available and could run the experiment for 16 days. The following experiment design was used:

| | Day | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cow | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 1 | | | | | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | | | | | |

All $2^4$ combinations occurred on each of the four cows. Effects AD, BC, and ABCD were completely confounded with days 1-4; effects AB, CD, and ABCD were confounded with days 5-8; effects ABD, AC, and BCD were confounded with days 9-12; and effects ABC, BD, and ACD were confounded with days 13-16. This design gave one-half information on the four factor interaction ABCD, and three-quarters information on the six two-factor and the four three-factor interactions. Full information was retained on the four main effects, A, B, C, and D. Measurements were taken at the start of a milking, in the middle, and at the end. Analyzing the data separately for these three times results in an analysis of variance as follows:

| | Degrees of freedom | Stage (mean square) Start | Middle | End |
|---|---|---|---|---|
| Total | 64 | | | |
| Correction for mean | 1 | | | |
| Days (ign. treat.) | 15 | | | |
| Cows | 3 | | | |
| Days x cows | 45 | | | |
| Treatment (elim. days) | 15 | | | |
| Main effects | 4 | | | |
| Interactions | 11 | | | |
| Remainder | 30 | | | |

For one stage the "interactions" mean square was 1.00, while that for the "remainder" was 8.00. If the interactions were truly zero, then this mean square of 1.00 should be an estimate of the error variance. It is highly unlikely that the mean squares of 8.00 and 1.00 are estimating the same parameter. There has to be something wrong with the model. If the original data were available, one could compute residuals using the above ANOVA model and try to determine what type of day x cow interaction is present. It is also possible that results from the midnight milking should have been omitted, since the experimenter was not present to oversee the conduct of the experiment.

## 5. Split Plot Design – An Example

In several instances, the very nature of animal experimentation dictates the use of a split plot design. The following is such an example. Thirty pregnant sows were used in a completely randomized design with ten sows being randomly allotted to each of three treatments, A, B, and C, designed to reduce the traumatic experience of birth. If a treatment can be found to achieve this, fewer piglets will die after birth. From each litter one randomly selected male and one randomly selected female piglet was obtained. The first born and the last born piglets were excluded. Measurements would be taken on the piglets at birth, at 12 hours after birth, at 24 hours, and at 36 hours. Birth weight would be used as a covariate.

By the very nature of the design, the sows are the whole plots, and the treatments A, B, and C are whole plot treatments. Sex is a split plot treatment. An analysis of variance table for each time is given below:

| Source of variation | Degrees of freedom | Mean squares: Birth | 12 hours | 24 hours | 36 hours |
|---|---|---|---|---|---|
| Total | 60 | | | | |
| Correction for mean | 1 | | | | |
| Treatment | 2 | | | | |
| Litters: treatments | 27 | | | | |
| Sex | 1 | | | | |
| Sex x treatment | 2 | | | | |
| Sex x sow: treatment | 27 | | | | |

Litters (sows) within treatments is the error for treatments, whereas sex x sow within treatment is the error for sex and sex x treatment interaction.

If the individual analyses do not suffice, and it is desired to combine the results, it is suggested that one use a multivariate analysis of variance with the variates being:

$Y_{0hij}$ = measurement at birth,

$Y_{12hij}$ = measurement at 12 hours after birth,

$Y_{24hij}$ = measurement at 24 hours after birth, and

$Y_{36hij}$ = measurement at 36 hours after birth.

One should use birth weight as a covariate and obtain adjusted sums of squares and products before completing the multivariate analysis.

## 6. Some Areas of Concern to the Consultant

Listed below are a number of areas which bother the author as a statistical consultant.

### (i) Unbalanced classification and linear model theory

Since the sample configuration is a random event, and perhaps related to some of the treatment or blocking variables, the analyses used are only conditional analyses, conditioned on the particular unbalanced configuration obtained. In a one-way classification, ignoring random sample sizes gives an over-estimate of the within variance component and an under-estimate of the between variance component.

### (ii) Covariance analyses

Textbooks universally consider the case of a common regression in an investigation. The assumption of a common regression is most likely an incorrect one in the majority of cases considered.

### (iii) Multivariate analysis

Heterogeneous variances and covariances are probably the rule rather than the exception. It is not known how robust the procedures are to this assumption being incorrect.

### (iv) Measurements

The idea of "test" animals and "nontest" animals in a pasture experiment can and has led to erroneous conclusions. In this procedure, additional animals are added to or subtracted from a pasture in order to keep the stocking rate constant. The animals left on the pasture throughout are called "test" animals, and their weight gains used in an analysis. One should use the total weight gain of all animals to assess the goodness of a pasture. Using only test animal weight gains could lead an experimenter to conclude that two pastures are equal, whereas, in truth, one was twice as good as the other if total weight gains were considered.

In nutrition experiments, total bone weight and bone ash weight are obtained. The experimenter then uses the ratio bone ash weight/total bone weight. Since the numerator and the denominator are highly correlated, the ratio is a constant

within-error fluctuation. Hence, any differences due to diets has been obliterated by taking the ratio. One should run an analysis on total bone weight. An analysis on bone ash weight would reveal little, if any, additional information. The analysis on the ratios is misleading and inappropriate.

(v) Outliers

How does one determine if an animal is sick or just an extreme deviate? When should animals be excluded? How valid is "being off their feed" a reason for excluding an animal? In a sheep nutrition experiment, if a ram tears off his horn but his weight gain, though down for one period, is about what the others gained for the whole period, should he be excluded or included?

It is suggested that one do residual analyses to look for outlying observations, outlying treatments, or an outlying block.

(vi) Data bases

Computer access and storage capabilities have lead people to consider collecting all kinds of data, regardless of their quality. Good informative observations are included in the same manner as non-informative bad data. Medical and veterinary records obtained from doctors or veterinarians are collected *en masse* and stored. Now, some individuals keep very accurate and precise records and others simply put in guesses. How does one establish a quality control procedure for data bases?

(vii) Animal breeding

Variance component quantitative genetics is the rule. Some or all of the assumptions for variance component usage are not met. This does not seem to bother the animal geneticist. Also, why don't they use major gene theory?

(viii) Recommended levels

The dosage level recommended may be too low as a result of too much dependence on tests of significance. For example, suppose that the response curve is as given in Figure 2, but the experimenter only observes values at $d_0$, $d_1$, $d_2$, $d_3$, and $d_4$. He then applies a significance test. Suppose that $d_0$ is significantly lower than $d_1$, $d_1$ is significantly lower than $d_2$, but none of the other differences are significant. Hence, $d_2$ is the recommended dose, whereas it should have been

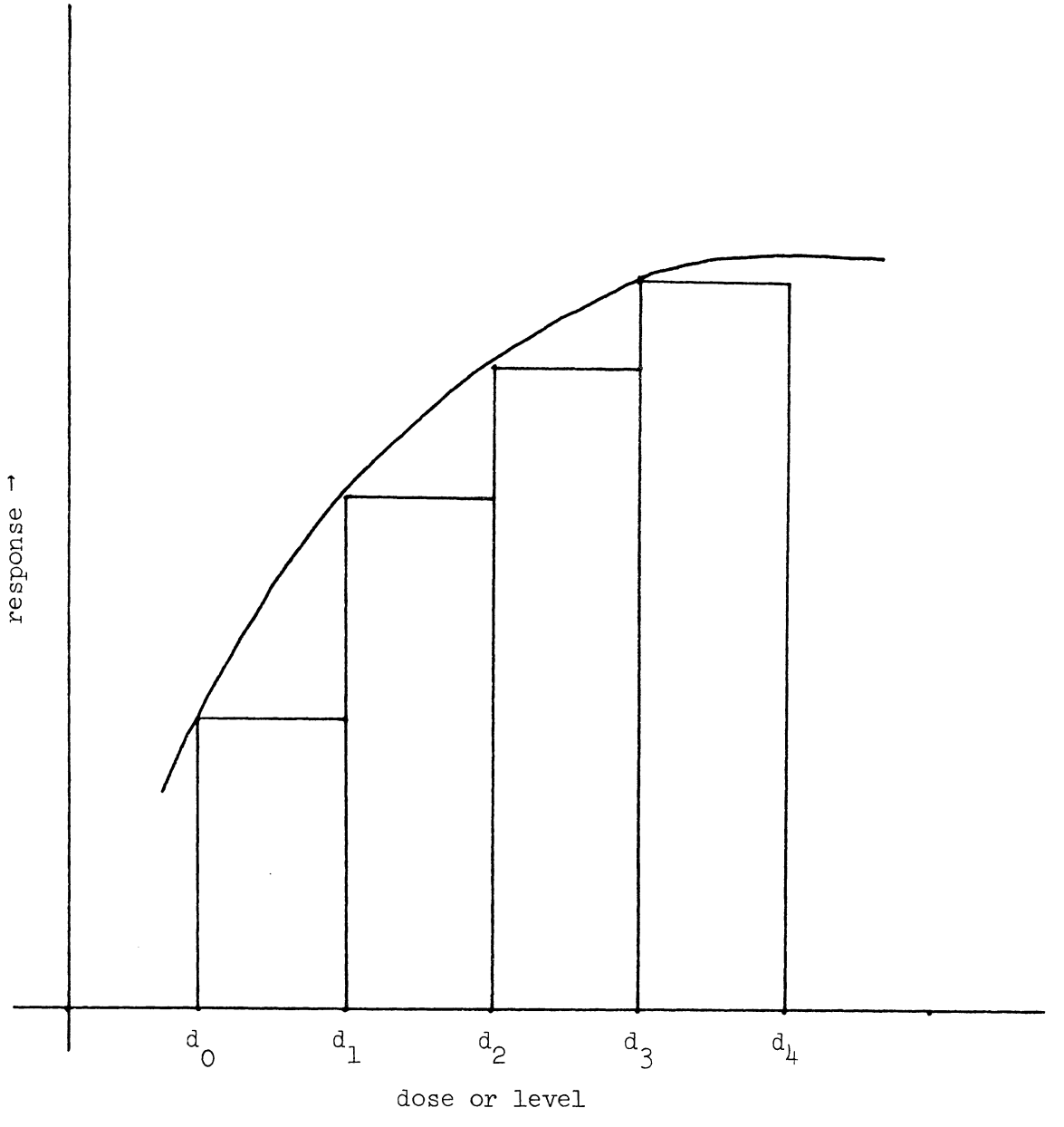


Figure 2. Response as a function of dose or level.

$d_4$. One should have fitted a response curve to the five data points and estimated a maximum.

In practice, the maximum response dose should probably not be recommended. A dose or level somewhat below the estimated maximum should be recommended to have an economic maximum. Also, in some cases, one would desire the least level that will produce the desired response. This is sometimes called a least effective dose.

(ix) Duncan's new multiple range and the Waller-Duncan test

It is used much more frequently than appropriate.

(x) Response plateaus

Suppose a response increases with level up to a certain point, and then plateaus with additional levels (see Figure 3). For example, if iron is added to the diet of laying hens, the egg shell will only become so hard. Adding additional iron to the diet does not increase the hardness of eggshells. In analyzing such data, should one use segmented regression (splines), or what should be done?

(xi) Lactation curves

A lactation curve using a single treatment might look like Figure 4a, which, in practice, would be truncated when the milk yield fell below an acceptable level. Now suppose that treatment A was given from zero to $d_0$, and then treatment B was used. B reduced milk yield (see 4b). A reduction in milk yield is easy. In 4c, suppose A was given from zero to $d_0$ and then treatment C was applied, causing an increase in production over what would have resulted from using A for the whole period (see dotted line). Now, how easy is it to have a cow perform as in 4c? It is easy to lower milk production, but difficult to raise it in the later stages of lactation. If the milk production can only be decreased, how should one apply treatments in a repeated measures design?

The above illustrate some of the problems encountered by the author as a statistical consultant. More questions have been raised than answers given. Much effort and study needs to be done in this area in order to be able to help the animal scientist, and to help the consultant in advising on the use of statistical procedures for such situations.

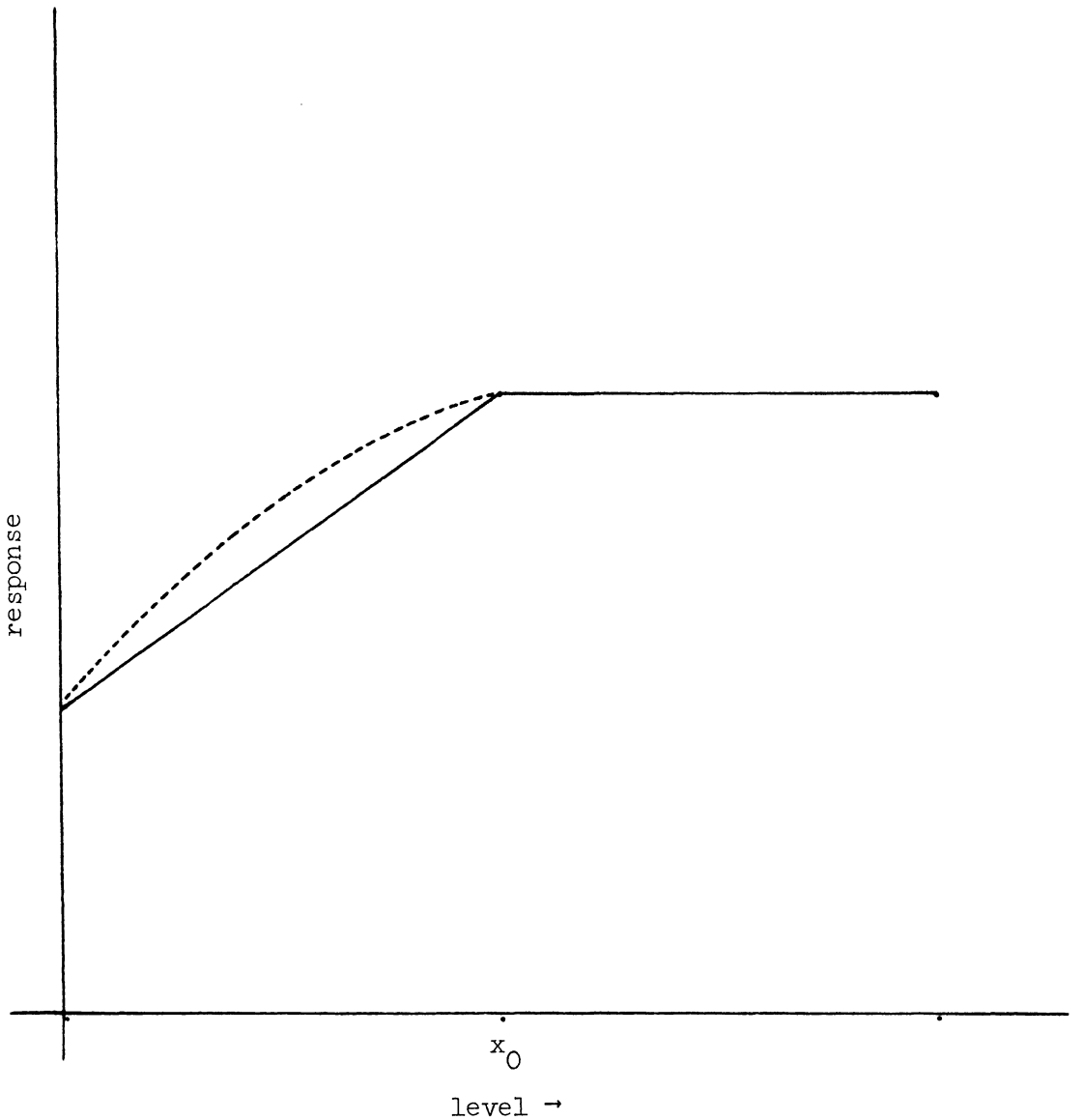


Figure 3. Response plateaus at $x_0$.

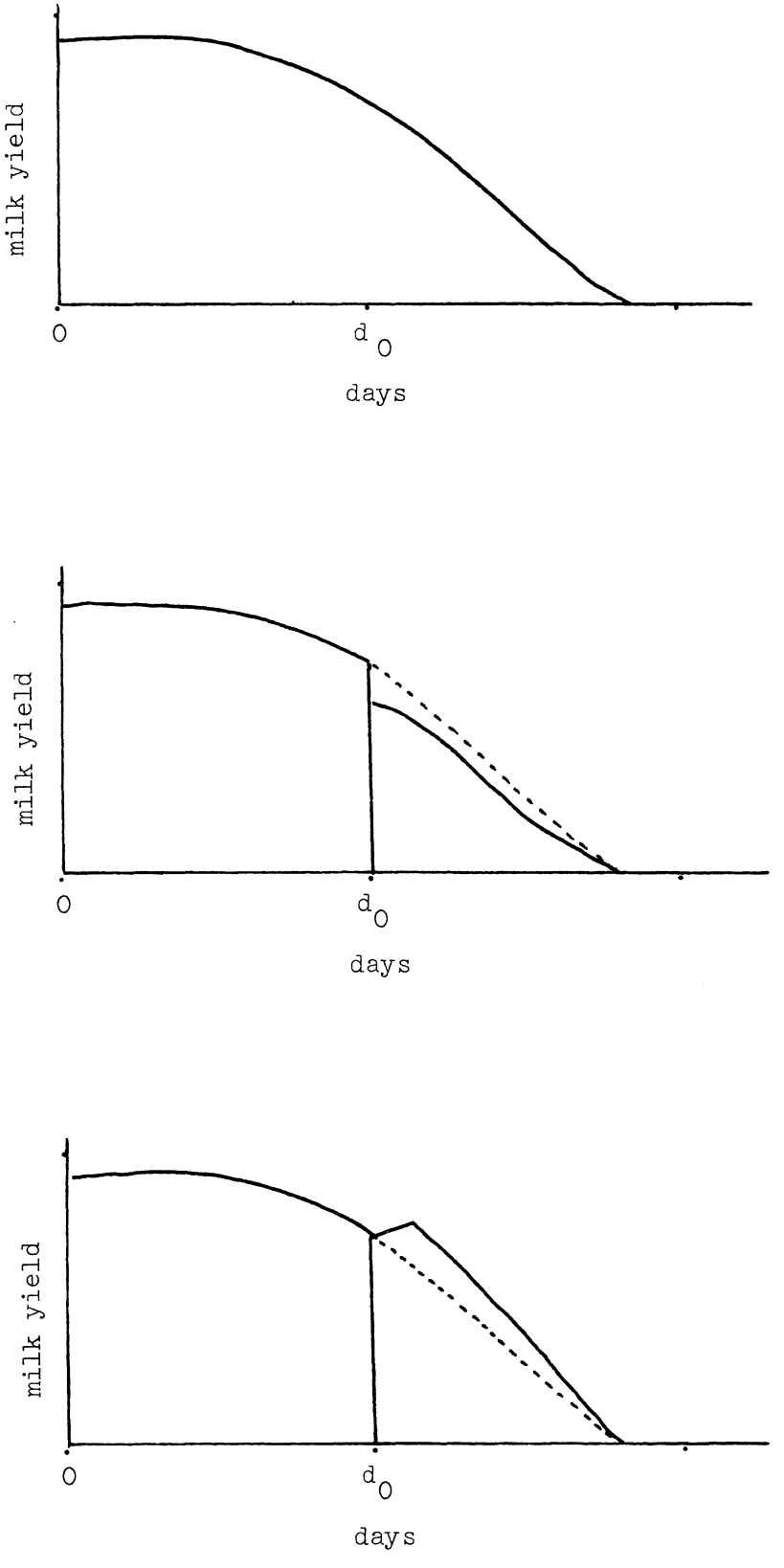


Figure 4. Lactation curves with intervention.

## Reference

Cox, C. P. (1958). The analysis of Latin square designs with individual curvatures in one direction. Journal of the Royal Statistical Society, B, 20:193-204.